

University of Groningen

## Estimation and prediction for nonlinear time series

Borovkova, Svetlana Alfredovna

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*

1998

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Borovkova, S. A. (1998). *Estimation and prediction for nonlinear time series*. [Thesis fully internal (DIV), University of Groningen]. s.n.

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

## 2 Review of correlation dimension estimators

Estimating the fractal dimension of a strange attractor from a corresponding time series has attracted considerable attention in the past few years and has become one of the main tools in the analysis of the underlying dynamics. Of all types of dimensions, most attention has been given to the *correlation dimension*. This is mainly because this type of dimension is easier to estimate than others and also because it provides a good measure of the complexity of the dynamics, i.e. of the number of active degrees of freedom (see Introduction).

Let  $(\mathcal{X}, T, \mu)$  be a dynamical system. Recall that the correlation dimension is defined via the correlation integral

$$C(r) = \mathbf{P}\{(X, Y) : \|X - Y\| \leq r\}.$$

where  $X, Y$  are independent, each having marginal distribution  $\mu$ . If there exists a constant  $\alpha$ , such that

$$C(r) \sim \text{const} \cdot r^\alpha \text{ as } r \longrightarrow 0, \quad (2.1)$$

then  $\alpha$  is called the correlation dimension of  $\mu$ . Note that

$$\alpha = \lim_{r \rightarrow 0} \frac{\log C(r)}{\log r}, \quad (2.2)$$

provided this limit exists.

The correlation dimension is a characteristic of the underlying invariant measure  $\mu$ . In a certain sense it characterises how smoothly  $\mu$  is distributed over the attractor: if  $\mu$  is a point measure, then  $\alpha = 0$ , and if  $\mu$  is absolutely continuous with respect to Lebesgue measure, then  $\alpha$  equals the topological dimension  $d$  of  $\mathcal{X}$ . These are two boundary cases, in general  $0 \leq \alpha \leq d$ .

A number of procedures for estimating the correlation dimension has been introduced in the literature. In the next sections we review some of these estimators and their properties.

## 2.1 Grassberger-Proccacia estimator

Grassberger and Proccacia [34] suggested a procedure of estimating  $\alpha$  which immediately became widely used by mathematicians and applied scientists. According to their method, we first estimate the correlation integral on the basis of the part of a stationary sequence of reconstruction vectors  $\{X_i\}_{i=1,\dots,n}$  by the sample correlation integral

$$C_n(r) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \mathbf{1}(\|X_i - X_j\| \leq r) \quad (2.3)$$

(note that  $C_n(r)$  is the proportion of pairs in the sample  $X_1, \dots, X_n$  no more than the distance  $r$  apart). For distance  $\|\cdot\|$  one usually takes the maximum norm, i.e. for a  $k$ -dimensional vector  $\mathbf{x}$  it is  $\|\mathbf{x}\| = \max_{1 \leq i \leq k} |x_i|$ . The estimate of  $\alpha$  is, however, independent of the choice of the norm (see [76]), and the max norm is taken as the most convenient one. The correlation integral  $C_n(r)$  is estimated for a vector of distances  $(r_1, \dots, r_l)$  and the Grassberger-Proccacia estimate for the correlation dimension  $\hat{\alpha}_n^{GP}$  is then obtained by studying the least-squares linear regression of  $\log C_n(r)$  vs.  $\log r$  (or, alternatively, the weighted regression).

In fact, the correlation dimension was initially introduced and defined by Grassberger and Proccacia via the sample correlation integral  $C_n(r)$  (and not  $C(r)$ ) as a double limit

$$\alpha^{GP} = \lim_{r \rightarrow 0+} \lim_{n \rightarrow \infty} \frac{\log C_n(r)}{\log r}. \quad (2.4)$$

The asymptotic properties of  $\hat{\alpha}_n^{GP}$  are related to the corresponding properties of the sample correlation integral. For stationary and ergodic sequences the a.s. consistency of the sample correlation integral follows from Theorem A of Aaronson *et al.* [1] (see Introduction), if for all  $r$  in the neighbourhood of 0 the following property holds:  $(F \times F)\{(x, y) : \|x - y\| = r\} = 0$ , where  $F$  is the distribution of  $X_1$ , and  $\|\cdot\|$  is the same norm as in (2.3). Moreover, if the dynamical system is weak Bernoulli, then it follows from

the result of Denker and Keller [26] that the asymptotic distribution of  $\sqrt{n}(C_n(r) - C(r))$  is normal with mean 0 and variance

$$\sigma^2 = 4\text{Var}[(h_1(X_1, r))] + 8 \sum_{k=1}^{\infty} \text{Cov}[(h_1(X_1, r)h_1(X_{k+1}, r))],$$

where

$$h_1(x, r) = \int_{\mathbf{R}^k} \mathbf{1}\{\|x - y\| \leq r\} dF(y).$$

From the variance of  $C_n(r)$  the asymptotic variance of the least squares estimator  $\hat{\alpha}_n^{GP}$  can be computed. Since  $\sqrt{n}(C_n(r) - C(r))$  is asymptotically normal, the distribution of  $[\sqrt{n}(C_n(r_i) - C(r_i))]_{i=1, \dots, l}$  approaches an  $l$ -dimensional normal distribution with mean 0 and covariance matrix  $[\sigma_{ij}]$  with components

$$\sigma_{ij} = 8 \sum_{k=1}^{\infty} \text{Cov}[h_1(X_1, r_i), h_1(X_{k+1}, r_j)] + 4\text{Cov}[h_1(X_1, r_i), h_1(X_1, r_j)].$$

Further, for a fixed choice of  $(r_1, \dots, r_l)$ , the vector  $[\sqrt{n}(\log C_n(r_i) - \log C(r_i))]_{i=1, \dots, l}$  also tends to an  $l$ -dimensional normal distribution with mean 0 and covariance matrix  $[a_{ij}]$ :

$$a_{ij} = \frac{\sigma_{ij}}{C(r_i)C(r_j)}.$$

Then  $\hat{\alpha}_n^{GP}$  is the least-squares estimator of  $\alpha$  where  $\sqrt{n}(\hat{\alpha}_n^{GP} - \alpha)$  is asymptotically normal with mean zero and variance

$$s^2 = \frac{1}{l^2} \sum_{i,j} b_i a_{ij} b_j, \quad (2.5)$$

where

$$b_i = \frac{\log r_i - \frac{1}{l} \sum_j \log r_j}{\frac{1}{l} \sum_j (\log r_j)^2 - (\frac{1}{l} \sum_j \log r_j)^2}.$$

The variance of  $C_n(r)$  and of  $\hat{\alpha}_n^{GP}$  can be consistently estimated from data by different methods, such as using  $U$ -statistics, Monte-Carlo simulation or the bootstrap. The variance estimation problem is very important for applications, since it allows us to compute confidence intervals for the estimates of the dimension. Chapter 6 is dedicated to this problem.

Note that the question of consistency of the Grassberger-Procaccia estimator makes sense only if in (2.1) the exact scaling holds, i.e. if, for some  $r_0 > 0$

$$C(r) = \text{const} \cdot r^\alpha \quad r \leq r_0. \quad (2.6)$$

If the exact scaling does not hold, we cannot talk about consistency of the Grassberger-Procaccia least-squares estimator, since the points  $r_i$ 's, where the correlation integral is estimated, do not approach 0, causing a non-vanishing bias in the estimate of  $\alpha$ . The asymptotic bias can be computed precisely and it is given by Cutler in [20]. She has also shown that the necessary and sufficient condition for the bias to be zero for any choice of  $(r_1, \dots, r_l)$ ,  $r_i \leq r_0$ , is the exact scaling (2.6) of the correlation integral, and, furthermore, if for some  $r > 0$  and  $s \in (0, 1)$  the vector of distances is taken to be

$$\mathbf{r}^{(l)} = (rs^l, rs^{l+1}, \dots, rs^{2l-1}), \quad (2.7)$$

then the asymptotic bias in the least-square estimate of the correlation dimension tends to 0 as  $l \rightarrow \infty$ .

From the results above it follows that, if the exact scaling (2.6) holds, then for stationary ergodic sequences  $\hat{\alpha}_n^{GP}$  is a strongly consistent estimator of  $\alpha$ , and, moreover, if the sequence is absolutely regular or is a functional of an absolutely regular process, then

$$\sqrt{n}(\hat{\alpha}_n^{GP} - \alpha) \xrightarrow{D} \mathcal{N}(0, s^2) \quad \text{as } n \rightarrow \infty, \quad (2.8)$$

where  $s^2$  is the asymptotic variance of  $\hat{\alpha}_n^{GP}$  given by (2.5).

Serinko [67] suggested an alternative approach to the least-squares estimator, which provides to a consistent (in a weak sense) estimator of the correlation dimension under the weak Bernoulli condition, even when no exact scaling holds. This estimator is based on the idea of introducing the sequence of distances  $\{r_n\}_{n \in \mathbb{N}}$  which tends to 0, but not too fast (for instance,  $r_n \rightarrow 0$ , but  $nr_n \rightarrow \infty$ ), and estimating the correlation integral at the points of this sequence. Then, by letting  $n \rightarrow \infty$ , we approach the double limit in (2.4) and achieve in this way consistency of this estimation procedure.

More precisely, Serinko considers the sequence  $\{r_n\}_{n \in \mathbb{N}}$  which converges to 0, but not faster than some other specified sequence  $\{b_n\}_{n \in \mathbb{N}}$ , i.e.  $r_n \rightarrow 0$ ,  $b_n \rightarrow 0$ , but  $r_n/b_n \rightarrow \infty$ , where the sequence  $b_n$  depends on the

mixing coefficients and other parameters. Then he shows that the following estimator

$$\hat{\alpha}_n = \frac{\log C_n(r_n)}{\log r_n}, \quad (2.9)$$

is *weakly consistent* if the dynamical system is weak Bernoulli, i.e. that

$$\hat{\alpha}_n \longrightarrow \alpha \text{ in probability, as } n \longrightarrow \infty,$$

under certain conditions on the mixing coefficients and on the dynamics.

Serinko also studied the combination of the least-square estimation procedure with his consistent approach. The idea is to eliminate the bias of the usual least-square estimator by letting both the sample size  $n$  and number  $l$  of points for regression tend to infinity simultaneously, i.e. by taking in (2.7) the sequence  $l_n \longrightarrow \infty$  as  $n \longrightarrow \infty$ . He specifies the rate with which the sequence  $l_n$  should tend to infinity by using the above result, which, together with Cutler's result on the asymptotic bias, leads to the weak consistency of the least-squares approach.

From a practical point of view, Serinko's estimator does not have an obvious advantage over the usual least-squares estimator, since the rate with which distances  $r_i^{(l_n)}$  should tend to 0 depends on unknown parameters, and has no clear practical value. Moreover, there is always a certain level of noise present in the system, and the estimates of the correlation integral at small distances are affected by it most, and thus, are the least reliable.

## 2.2 Takens estimator

Takens proposed an alternative approach of estimating the correlation dimension [77]. Assuming again that the exact scaling (2.6) holds, Takens first considered estimating  $\alpha$  from i.i.d. realisations  $R_i = \|X_i - Y_i\|$  of the distance  $\|X - Y\|$ , where  $X_i$  and  $Y_i$  are independent each having the marginal distribution  $\mu$ . In the case of (2.6), the conditional distribution of  $U_i := R_i/r_0$  given  $R_i \leq r_0$  is

$$\mathbf{P}(U_i \leq t | U_i \leq 1) = t^\alpha, \text{ for } t \in [0, 1].$$

Then the distribution of  $S_i = -\log U_i$  is exponential with parameter  $\alpha$ , i.e. the distribution function of  $S_i$  is

$$G(s) = \mathbf{P}(S_i \leq s) = 1 - e^{-\alpha s}$$

and the density is

$$g(s) = \alpha e^{-\alpha s} \cdot \mathbf{1}_{[0, \infty)}(s).$$

If we are given an i.i.d. sample  $S_1, S_2, \dots, S_N$  from the distribution  $G(s)$ , then the Maximum Likelihood estimator (MLE) for the parameter of exponential distribution is given by

$$\hat{\alpha}^{ML} = \frac{N}{\sum_{i=1}^N S_i} \quad (2.10)$$

and the Uniform Minimal Variance Unbiased Estimator (UMVUE) of  $\alpha$  is

$$\hat{\alpha} = \frac{N-1}{\sum_{i=1}^N S_i},$$

since it is unbiased and is a function of the complete sufficient statistics  $\sum_{i=1}^N S_i$ . Both estimators are strongly consistent when  $S_i$ 's are i.i.d.

In general, independent realisations of the distances  $\|X - Y\|$  will not be available (moreover,  $X_i$ 's themselves are, in most cases, not independent) and thus a modification of the estimator (2.10) may become necessary. Given a finite segment  $X_1, \dots, X_n$  of a stationary sequence of the reconstruction vectors, we can form  $n(n-1)/2$  pairwise distances  $\|X_i - X_j\|$ . Takens suggested to use the estimator

$$\hat{\alpha}_n^T = - \left( \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \log \frac{\|X_i - X_j\|}{r_0} \right)^{-1}. \quad (2.11)$$

Note that (2.11) is the reciprocal of the  $U$ -statistic with  $m = 2$  and the kernel function  $h(x, y) = \log \|x - y\|$ . However, Theorem A cannot be applied directly, since  $h(x, y)$  is in general not bounded. It turns out that there are examples of stationary ergodic and mixing sequences for which the Takens estimator is not consistent. In Chapter 3 we will address this question in more detail and will show consistency of the Takens estimator under some additional conditions.

In fact, an estimator similar to (2.11) was first introduced by Hill [40] in the context of estimating the tail index of a distribution: suppose we have a random sample  $T_1, T_2, \dots, T_N$  from a distribution  $F$  which behaves at 0 as

$$F(x) \sim \text{const} \cdot x^\alpha \text{ as } x \rightarrow 0,$$

and we want to estimate  $\alpha$  without making assumptions about the form of  $F$  elsewhere. Assuming that actually  $F(x) = Cx^\alpha$  when  $x \leq x_0$ , for some known  $x_0$  and some constant  $C$ , the conditional Maximum Likelihood estimator of  $\alpha$  (Hill's estimator) is given by

$$\hat{\alpha}^H = - \left( \frac{1}{m} \sum_{i=1}^m \log \frac{T_{(i)}}{T_{(m+1)}} \right)^{-1}, \quad (2.12)$$

where  $T_{(1)} \leq T_{(2)} \leq \dots \leq T_{(m)} \leq T_{(m+1)} \leq x_0$  are the order statistics which are below the threshold  $x_0$ . Note that the estimators (2.11) and (2.12) coincide up to a scaling factor: in Hill's estimator observations are scaled by the last order statistics which is still below the threshold, and in Takens estimator by the threshold itself.

The strong consistency and asymptotic normality of Hill's estimator for i.i.d. observations was shown already by Hill [40]. For stationary mixing sequences the analogous result was shown by Hsing [43], under certain conditions on the distribution  $F$  and the dependence structure of the process.

## 2.3 Other estimators

**Chord estimator** The chord estimator was suggested by Theiler and Lookman [80] and it makes use of an “easy ” way of fitting a straight line  $\log C(r)$  vs.  $\log r$ . Their argument was that the easiest and the most obvious way to estimate a slope of the correlation integral is to choose two distances  $r_0$  and  $r_1$ , estimate the correlation integral in these two points by the sample correlation integral and measure the slope of the chord that is drawn through those two points.

Then formally the chord estimator is

$$\hat{\alpha}^C = \frac{\log C_n(r_0) - \log C_n(r_1)}{\log r_0 - \log r_1}. \quad (2.13)$$

This estimator seems at first sight very inefficient, since it throws away all the information except for the distances  $r_0$  and  $r_1$ . However, Theiler and Lookman have shown that, under the hypothesis of independence of the distances  $\rho_{ij} = \|X_i - X_j\|$ , this estimator is only a factor 1.25 worse (in terms of a relative error) than the Takens estimator, which uses *all*



the distances below a certain threshold. And, of course, the chord estimator is much easier and faster to compute than any other estimator of the correlation dimension.

The main question when using this estimator is the choice of the optimal chord (i.e. of two distances  $r_0$  and  $r_1$ ). Choosing the chord has two aspects: on the one hand, we may want to choose the chord as long as possible, since the effect of fluctuations on the endpoints is less noticeable in that case. On the other hand, if we take the chord that is too long, then there are two aspects to consider. First, the left point  $r_0$  would be too small, i.e. we would have to consider distances that are too small, where we usually do not have enough observations and thus fluctuations are particularly high. Second, the chord may not lie anymore in the area where the scaling rule (2.1) holds.

Theiler and Lookman deduced the optimal rule for choosing the chord by taking the chord which minimises the relative variance of this estimator. Assuming the exact relationship (2.6), the relative variance has a form

$$\frac{\sigma^2}{\alpha^2} = \frac{\Theta - 1}{n_0(\log \Theta)^2}, \quad (2.14)$$

where  $\Theta$  is a “length” of a chord in terms of a number of distances, i.e.

$$\Theta = \frac{n_0}{n_1} = \frac{C(r_0)}{C(r_1)} \approx \frac{C_n(r_0)}{C_n(r_1)},$$

and  $n_0$  (resp.  $n_1$ ) is the number of distances smaller than the upper cut-off point  $r_0$  (resp. lower cut-off point  $r_1$ ). The expression (2.14) is minimised for  $\Theta \approx 5$ , so, having chosen the upper distance  $r_0$ , the optimal lower distance should be taken such that

$$\frac{C_n(r_0)}{C_n(r_1)} = \Theta_{\text{optimal}} = 5.$$

This rule gives a relative error for  $\alpha$  about  $1.25/\sqrt{n_0}$ , while the Takens estimator has a relative error  $1/\sqrt{n_0}$ .

All the above computations are made assuming that the distances are independent, which, as we already noted, is in general not the case. However, the estimator for the correlation integral  $C_n(r)$  is strongly consistent for stationary ergodic sequences, so, if we assume, as in the case of Takens estimator, that for  $r \leq r_0$  the exact scaling rule (2.6) holds, then the

chord estimator is consistent. Moreover, its variance can be computed via the variance of the sample correlation integral, in the same way as in the previous section.

**Ellner estimator** This estimator, suggested by Ellner [33], combines the ideas of both the Takens estimator and the chord estimator. Here we also chose 2 cut-off distances: the upper  $r_0$  and the lower  $r_1$ , and base our estimate on all the distances between  $r_1$  and  $r_0$ . Assuming again that for  $r \in [r_1, r_0]$  the exact scaling (2.6) holds, the Maximum Likelihood estimator of  $\alpha$  is

$$\hat{\alpha}^E = - \left( \frac{n_1}{n_0 - n_1} \log\left(\frac{r_1}{r_0}\right) + \frac{1}{n_0 - n_1} \sum \log \frac{\rho_{ij}}{r_0} \right)^{-1}, \quad (2.15)$$

where  $n_0$  (resp.  $n_1$ ) is a number of distances smaller than  $r_0$  (resp.  $r_1$ ) and the sum extends over all distances  $\rho_{ij}$  between  $r_0$  and  $r_1$ . Note that as  $r_1 \rightarrow 0$  we obtain the original Takens estimator.

Ellner has shown (under the assumption of independence) that the relative variance of this estimator is given by

$$\frac{\sigma^2}{\alpha^2} = \frac{1}{n_0[1 - (r_1/r_0)^\alpha]}. \quad (2.16)$$

In contrast with Takens estimator, this estimator *is* strongly consistent for stationary ergodic sequences, provided  $C(r) = C \cdot r^\alpha$  for  $r \in [r_1, r_0]$ , since, due to cutting off all distances less than  $r_1$ , the problem of unboundedness of  $\log \|X_i - X_j\|$  does not occur and, hence, Theorem A can be applied directly.